Fergusson DM, Horwood LJ. Transitions to cigarette smoking during adolescence. *Addictive Behaviors*, 1995; 20(5): 627-642.

Transitions to Cigarette Smoking During Adolescence

David M Fergusson and L John Horwood

Department of Psychological Medicine

Christchurch School of Medicine, Christchurch, New Zealand

Running Head: SMOKING IN ADOLESCENCE

Correspondence: Associate Professor David M Fergusson

Christchurch Health and Development Study

Christchurch School of Medicine,

CHRISTCHURCH

New Zealand.

Abstract

The process of transition from non-smoking to regular weekly smoking during the period from 10 to 16 years was examined using data gathered during the course of a longitudinal study of 957 New Zealand adolescents. These data were analyzed using a latent Markov model to estimate both rates of transition between stages of smoking and errors of measurement in report data. The fitted model suggested that the process of transition to smoking was a progressive and one way process in which once teenagers had graduated to a given smoking status return to earlier stages was uncommon. This process also showed a clear tendency to accelerate with age so that rates of transition to smoking behaviors after the age of 14 years were far higher than rates of transition at age 10 years. The model also made it possible to estimate the accuracy of smoking reports. This suggested that individuals who were non-smokers or regular smokers were classified with better than 95% accuracy on the basis of report data. However, the reporting accuracy of occasional smoking was poor with 42% of occasional smokers being falsely classified as non-smokers. The implications of these findings for the understanding of the processes of transition to smoking behaviors and the effects of measurement errors in report data are discussed.

There has been a large amount of research into the processes and risk factors that lead to the development of cigarette smoking in childhood and adolescence (for reviews see Conrad, Flay & Hill, 1992; Holland & Fitzsimmons, 1991; Kandel, 1980; Miller & Slap, 1989; Moncher, Holden & Schinke, 1991). One aspect of this research has focussed on the ways in which smoking experiences and experimentation in childhood lead to later regular cigarette smoking in adolescence and young adulthood. This research has suggested the presence of clear linkages between patterns of early experimentation with cigarettes and later regular smoking behaviors with children who begin smoking or smoking experimentation in early or middle childhood having increased risks of later smoking (Chassin & Presson, 1990; Escobedo, Marcus, Holtzman & Giovino, 1993; Krohn, Massey, Skinner & Lauer, 1983; Taioli & Wynder, 1991). However, whilst the linkages between early smoking behaviors and later regular cigarette smoking have been well documented, less is known about the processes by which children make transitions from early smoking experimentation to later smoking and, in particular, detailed longitudinal data describing the processes by which smoking behaviors are acquired, develop or show remission, is limited.

In this paper we use data from a six year longitudinal study of the development of cigarette smoking behaviors in a sample of New Zealand children studied from ages 10 to 16 years to develop a latent Markov model of the processes of transition which led to the development of regular smoking behaviors in adolescence. The theoretical and statistical background to this analysis is developed below.

Theoretical and Statistical Background

To motivate the analysis consider some sample or population of children studied at a series of time periods t_1, t_2 .... t_n with measures of smoking behaviors being observed at each time. For simplicity, assume that at each time period children are classified as smokers or non-smokers using some standardized criterion of smoking behaviors. For a series of n time periods there are potentially 2^n response patterns that describe the history of smoking and non-smoking over the n time periods and the task of describing the development and change in smoking behaviors over time amounts to developing an account of the ways in which subjects make transitions between the states "smoker" and "non-smoker" over the n time periods.

One approach to charting changes in state over time is to devise transition matrices that describe the probability that a subject who has a particular smoking status at a given time will be observed to be a smoker or non-smoker at a later time. An example of the state to state transition matrix is shown below.

		t + 1	
		Non-Smoker	Smoker
t	Non-Smoker	a	1 - a
	Smoker	1 - b	b

The transition matrix shows the conditional probability that a subject who is in a given state at some time t will remain in this state or change state at some later time t + 1. The parameters of this matrix may be interpreted in the following ways:

1. The parameter "a" describes the probability that a subject who is a non-smoker at time t will remain a non-smoker at time t + 1. Accordingly, "1 - a" gives the probability that subjects who are non-smokers will show an onset of smoking in the interval from t to t + 1.

2. The parameter "b" shows the probability that a subject who is a smoker at time t will remain a smoker at time t + 1. Accordingly, "1 - b" represents the probability that an individual who was a smoker at time t will show remission or cessation of smoking at time t + 1.

By developing empirical transition matrices is becomes possible to chart the ways in which smoking behaviors change over time and, in particular, to examine the onset and remission of smoking behaviors. This strategy has implicitly underwritten studies that have attempted to examine the linkages between early smoking behaviors and later outcomes. However, the empirical Markov model described above suffers from one major limitation to the extent that it assumes the classification of the subject's status at different times is made without error. In the case of smoking behaviors, this assumption is clearly unrealistic since it is well known that measures and reports of smoking behaviors are subject to what may be quite substantial errors of measurement (Fergusson & Horwood, 1989; Gillies, 1985; Gillies, Wilcox, Coates, Kristmundsdottir & Reid, 1982; McKennell, 1980; Stacy, Flay, Sussman, Brown, Santi & Best, 1990). In turn, errors of measurement may lead to compromised estimates of patterns of change in smoking behavior since with data containing errors changes in status can be attributed to: a) true changes in smoking behavior over time; b) apparent changes due to measurement error. These considerations suggest that to adequately describe transitions in smoking behavior it is necessary to develop methods which take account of possible errors of measurement in the classification of smoking behaviors.

One approach to minimizing measurement errors in reports of smoking has been through the use of various empirical methods designed to improve reporting accuracy or the detection of individuals who smoke. These methods include the use of carefully designed and standardized questionnaires (Gillies, 1985), the use of biochemical measures (Gillies et al., 1982; Luepker, Pechaeck, Murray, Johnson, Hurd & Jacobs, 1981; McNeil, Jarvis, West, Russell & Bryant, 1987) or other techniques including the so called "bogus pipeline" (Murray & Perry, 1987). All of these methods have some liabilities: even carefully designed questionnaires may lead to false positive and false negative responses (Fergusson & Horwood, 1989; Gillies, 1985), biochemical measures fail to provide estimates of cigarette intake and may be unreliable for individuals in the early stages of smoking or who smoke infrequently (McNeil et al., 1987) and there are ethical problems with techniques such as the "bogus pipeline". An alternative approach to addressing errors of measurement in reports of substance use behaviors has been provided by the use of latent variable modelling methods (Fergusson & Horwood, 1989; Newcomb & Bentler, 1988). These methods provide techniques for estimating the magnitude of errors of measurement in report data and thence of providing estimates of the associations between variables corrected for errors of measurement.

To address problems of measurement error in Markov models, latent Markov models have been devised. These models were first proposed by Wiggins (1973) and have recently been refined in the work of Van de Pol and his associates (Langeheine & Van de Pol, 1990; Van de Pol & Langeheine, 1990; Van de Pol & de Leeuw, 1986). The essential difference between the latent Markov model and the empirical Markov model is that the latent Markov model draws a theoretical distinction between:

i) The subject's true but non-observed changes in status over time.

ii) Observed changes in status.

The general aims of the model are to secure (subject to certain assumptions), estimates of the true processes of change occurring over time and estimates of the extent of measurement error in the observed data. The approach is best introduced by way of an example.

The Table below shows the empirical Markov model described earlier reformulated as a latent Markov model. This specification may be interpreted as follows:

1. The transition matrix represents the state to state transitions occurring in the subjects' true but non-observed smoking behaviors at times t, t + 1. In this matrix the parameter a' represents the probability that a child who was a non-smoker at time t would remain a non-smoker at time t + 1 and the parameter b' represents the probability that a child who was a smoker at time t would remain a smoker at time t + 1.

2. The model recognizes the distinction between the child's true status represented by the latent state and the observed status based on report or similar data by introducing a response vector which represents the probability that a subject who is in a given latent state will be classified as a smoker. The parameter g_1 represents the probability that a child who is a non-smoker will be falsely classified as a smoker on the basis of observed data and the parameter g_2 represents the probability that a child who is a smoker will be correctly classified as a smoker on the basis of observed data. The probabilities of non-smoking conditional on latent status are given by the complements: $1 - g_1$; $1 - g_2$.

The Latent Markov Model

a) Latent Transition Matrix

t + 1

Non-Smoker Smoker

t	Non-Smoker	a'	1 - a'
	Smoker	1 - b'	b'

b) Response Vector

		Observed State	
		Pr(Smoker)	Pr(Non-Smoker)
Latent State	Non-Smoker	g_1	$1 - g_1$
	Smoker	g_2	$1 - g_2$

The latent Markov model assumes that the processes that generate change in the subject's smoking status over time conform to a first order Markov process in which the subject's history of changes up to time t is summarized by his/her status at time t. This implies that the only information required to predict status at time t + 1 is the subject's status at the previous point of observation. Subject to the availability of at least three times of observation, the parameters of this model can be estimated from the subject's observed history of smoking behaviors using methods of maximum likelihood estimation (Van de Pol, Langeheine & de Jong, 1991). These methods estimate the model parameters (a', b', g_1, g_2) by maximizing the likelihood of observing the data conditional on the set of model parameters. (Readers unfamiliar with maximum likelihood estimation may find it easiest to conceptualize this process by noting that the maximum likelihood estimates are approximately equivalent to finding parameters which minimize the Pearson chi square goodness of fit value between the observed data and the data implied by the model parameters). Additionally such models have the feature of being falsifiable to the extent that the number of model parameters to be estimated will typically be smaller than the number of observed response patterns and this makes it possible to ascertain the goodness of fit between the observed data and that implied by the model parameters. This leads to a logic of falsification in which poorly fitting models can be rejected as inadequate accounts of the data, whilst well fitting models show that the data are generally consistent with a given set of model assumptions (Goodman, 1974).

Thus far we have illustrated the ideas of empirical and latent Markov models using a simple two state (smoker, non-smoker) model. However, in describing smoking during childhood this conceptualization is likely to be inadequate since children are likely to pass through a number of stages in the development of smoking behaviors with these stages ranging from non-smoker, through experimental or occasional smoking to regular smoking. A further complication of the study of childhood smoking behaviors is that the rates at which children make transitions between stages of smoking may vary with age and also the reporting accuracy of smoking behaviors may vary with age. Collectively, these considerations suggest that useful accounts of the development of cigarette smoking during childhood are likely to be characterized by the following features:

i) The child's smoking status is classified according to the extent to which the child smokes.

ii) That errors of measurement in reports of smoking are taken into account using a latent Markov (or similar) model.

iii) That the model permits time dependency in the process studied by allowing both state to state transitions to vary with age and measurement accuracy to vary with age.

In this paper we report on the results of fitting a three state (non-smoker, occasional smoker, regular smoker) model to data gathered on reports of smoking behaviors in a birth cohort of New Zealand children studied to the age of 16 years. This model uses a latent Markov formulation which permits: a) the estimation of measurement errors in report data and estimation of rates of transition corrected for measurement error; b) possible time dependency in rates of transition with age; c) possible time dependencies in reporting errors with age. The general aims of the analysis were to provide an account of the processes by which members of this cohort made transitions from being non-smokers to being occasional or regular smokers during the period from middle childhood into adolescence.

Method

The data described in this paper were gathered during the course of the Christchurch Health and Development Study. The Christchurch Health and Development Study is a longitudinal study of a birth cohort of 1265 children born in the Christchurch urban region during mid 1977. These children have been studied at birth, four months and annual intervals to the age of 16 years using data collected from a variety of sources including parental interviews, self report, teacher questionnaires and information gathered from medical records. An account of the design of the research has been given previously (Fergusson, Horwood, Shannon & Lawton, 1989). The data analyzed here are based on measures of cigarette smoking behaviors observed at two yearly intervals from the age of 10 to 16 years. A discussion of the method of classifying smoking behaviors is given below.

At 10, 12, 14 and 16 years reports of child smoking behaviors were obtained from parallel self and parent reports with these reports being collected during separate interviews conducted at different sites: parents were interviewed at home, whilst children were interviewed at school. At each interview parents and children were asked to rate the frequency of cigarette smoking by the child on a five point scale ranging from "does not smoke" to "smokes at least once a week". In this questioning descriptions of the child's smoking behavior at around the time of interview were obtained but an exact timeframe for measuring smoking behaviors was not specified. For individuals who were smokers, estimates of daily cigarette intake were also obtained. For the purposes of the present analysis parent and child data were combined to give a best estimate of the child's likely smoking behaviors reported by two sources.

It was clear from the report data that smoking behaviors in childhood varied more or less continuously from children who were committed non-smokers to those who were committed and addicted smokers with other children lying between these extremes. To present this variability using a latent Markov model it was necessary for this continuum of smoking behaviors to be quantilised into a series of discrete groups with each group representing a relatively homogeneous set of individuals showing similar levels of smoking behavior. The decisions for quantilising the distribution of smoking into a series of discrete stages were based on two considerations: a) *a priori* assumptions about the classification of children into theoretically meaningful groups; b) examination of the distribution of responses to ensure that any classification contained sufficient observations to meet statistical and analytical demands. These considerations suggested that the most useful classification of the sample was into three groups:

i) Non-smokers. Children who were described as non-smokers on the basis of both parental and self report.

ii) Occasional smokers. Children who were classified as smoking on the basis of either parental or self report but who smoked less than once per week.

iii) Regular smokers. Children who were classified as smoking on the basis of either parental or self report and who reported smoking at least once a week.

It would have clearly been desirable to further subdivide the regular smokers into those who engaged in daily smoking and those who smoked less frequently. However, the numbers of children engaging in daily smoking were not sufficiently large prior to 16 years for this classification to yield adequate numbers of subjects for analysis. The model proposed thus describes the processes by which children made transitions from being non-smokers to being fairly regular smokers by the age of 16 years.

Sample

The analyses in this report are based on a sample of 957 children. This sample represents 76% of the original birth cohort of 1265 children enrolled in the Study and 86% of the sample members resident in New Zealand at age 16 years.

While these sample retention rates are high for a longitudinal study it is possible that sample loss processes could have produced a non-random sampling of the original cohort. To examine this possibility the obtained sample of 957 children was compared with the remaining 308 cohort members not included in the analysis on a range of socio-demographic measures collected at the time of the initial (birth) interview. This analysis showed no significant differences between the two samples with respect to: maternal age ($p>.15$), parental smoking ($p>.05$), gender ($p>.40$), ethnicity ($p>.80$) or family size ($p>.30$). However, there were small but detectable tendencies for the obtained sample to be under represented by children from single parent families ($p<.01$), children from families of lower socio-economic status ($p<.01$), and children whose mothers lacked formal educational qualifications ($p<.05$), suggesting a slight tendency for children from more socially disadvantaged backgrounds to be excluded from the sample.

At the same time, while the analysis detected some departures from simple random sampling assumptions, any bias in the sample was relatively small and it is doubtful that this differential sample loss would have materially influenced the parameter estimates reported later. Some reassurance that this is the case is provided by a previous analysis in which associations between child smoking and other factors were corrected for sample selection bias using the method described by Berk (1983). This analysis suggested that differential sample losses had no detectable effect on the estimation of parameters including estimates of the association between early smoking behaviors and later smoking (Fergusson, Lynskey & Horwood, 1994).

Results

Observed Transitions in Cigarette Smoking (10-16 Years)

Table 1 gives a summary of the transitions between stages of cigarette smoking made by the sample over the period from 10 to 16 years. The Table shows for each age (10, 12, 14, 16 years) the sample classified into three groups - non-smoker, occasional smoker and regular smoker - using the criteria described earlier. For each adjacent pair of ages the Table shows the observed transition probabilities that a subject who had a given smoking status at some age would have a given status two years later. The Table leads to the following conclusions:

1. During the period from 10 to 12 years, of those who were non-smokers at age 10: 90% remained non-smokers two years later; 10% became occasional smokers and less than 1% became regular smokers. Amongst those who were occasional smokers at age 10 years: 56% returned to being non-smokers, 41% remained as occasional smokers and 3% became regular smokers. Amongst

those who were regular smokers at 10 all but one remained either occasional or regular smokers at age 12 years.

2. During the period from 12-14 years, of those who were non-smokers at age 12: 83% remained non-smokers two years later; 14% became occasional smokers and 3% became regular smokers. Amongst those who were occasional smokers at age 12 years: 47% returned to being non-smokers; 35% remained occasional smokers and 18% became regular smokers. For those who were regular smokers at age 12 the majority (79%) remained either occasional or regular smokers.

3. During the period from 14-16 years, 67% of those who were non-smokers at age 14 remained non-smokers two years later; 20% became occasional smokers and 13% became regular smokers. Amongst occasional smokers at age 14, 28% became non-smokers; 26% remained occasional smokers and 45% became regular smokers. Of the regular smokers at age 14 the great majority (91%) remained regular smokers at age 16 years.

Inspection of the process of transition to smoking shows evidence of a clear acceleration with age. This is evident clearly in the transition rates from non-smoking to smoking behaviors: during the period from 10-12 years 90% of those who were non-smokers remained non-smokers two years later but during the period from 14-16 only 67% of those who were non-smokers remained non-smokers two years later.

INSERT TABLE 1. HERE

Whilst the data in Table 1 provides an account of the process of transition between various smoking stages, the interpretation of these data is complicated by the fact that the observed data were subject to errors of measurement and as a result the observed state to state transitions may give a misleading impression of the process of transition occurring within the sample. To address this issue the latent Markov model described earlier in this paper was fitted to the data.

Model Fitting

To examine the model most suitable to describe the observed data, a series of nested models was fitted. These models varied in the extent to which it was assumed that: a) transitions between latent states varied with time of measurement and b) reporting errors varied with time of measurement. Specifically, the models fitted were:

1. The Fully Constrained Model (Model 1). This model assumed that transitions between smoking states over time did not vary with the subject's age and that the same latent matrix described all transitions. The model also assumed that reporting accuracy was constant at all ages.

2. The Time Dependent Markov Model (Model 2). This model assumed that the latent transition matrices varied with the time of observation so that the transition matrix describing the period from 10-12 (for example) was different from the transition matrix describing the period from 14-16 years (for example). The model, however, assumed that reporting accuracy remained constant at all ages.

3. The Time Dependent Markov Model with Time Dependent Errors (Model 3). The final model fitted assumed that both the latent transition matrix and errors of measurement varied with the time of observation. To identify this model it was assumed that: a) errors of measurement at ages 10 and 12 were described by a common response vector; b) errors of measurement at ages 14 and 16 were described by a common response vector; c) that the response vector describing ages 10 and 12 could differ from the response vector describing ages 14 and 16 years.

Table 2 reports on tests of the adequacy of these three models. This Table shows for each model the log likelihood ratio chi square test of the goodness of fit of the model to the observed data and tests of the relative improvement of fit of the series of nested models obtained by taking differences in the goodness of fit values. This Table leads to a generally clear set of conclusions about the best fitting and most parsimonious account of the data. First the Table shows that the fully constrained model (Model 1) did not fit the observed data adequately ($LR\chi^2 = 158.9$; df = 66; $p<.0001$) suggesting the presence of time dependent processes in either the latent transition matrices or the response vectors. The model permitting time dependent transition matrices but constant errors (Model 2) fitted the data well ($LR\chi^2 = 31.9$; df = 54; $p>.99$) as did the model assuming time dependence of both transition matrices and measurement errors ($LR\chi^2 = 26.3$; df = 48; $p>.99$). However, the model assuming time dependence of both transition matrices and measurement errors (Model 3) did not lead to a significant improvement in fit ($LR\chi^2 = 5.6$; df = 6;

p>.25) over the model assuming time dependent transition matrices and constant measurement errors. These comparisons led to the conclusion that the best fitting and most parsimonious account of the data was one in which:

i) Rates of transition between latent smoking states varied with age.

ii) Errors of measurement in reports of smoking behaviors were similar at different ages.

INSERT TABLE 2. HERE

Table 3 shows the estimated parameters of the fitted Model 2. This Table gives:

1. Estimates of the Initial Vector for the Model. This shows estimates of the proportion of subjects in each of the latent states (non-smoker, occasional smoker, regular smoker) at the first point of observation (10 years). These estimates suggest that 86% of the sample were non-smokers at age 10 years, 13% were occasional smokers and less than 1% were regular smokers.

2. Estimates of the Latent Transition Matrices. These show the probabilities of state to state transitions between various stages of smoking taking into account reporting errors (subject to the general model assumptions). These results show clear evidence of time dependence in the process of transition to smoking.

During the period from 10-12 years, 93% of non-smokers at age ten years remained non-smokers; 7% became occasional smokers and less than 1% became regular smokers. Of those who were occasional smokers at age 10 years, 98% remained occasional smokers and 2% became regular smokers. Of those who were regular smokers at age 10 years all remained regular smokers at age 12 years.

During the period from 12-14, there was clear evidence of a tendency for transitions to smoking behaviors to increase. Of those who were non-smokers at age 12 years, 86% remained non-smokers; 12% became occasional smokers and 2% became regular smokers. Of those who were occasional smokers, 74% remained occasional smokers 7% became non-smokers and 19% became regular smokers. Of those who were regular smokers at age 12, all remained regular smokers at age 14 years.

During the period from 14-16, rates of transition to smoking showed a further tendency to accelerate. Of those who were non-smokers at the age of 14 years, 62% remained non-smokers; 29% became occasional smokers and 9% became regular smokers. Of those who were occasional smokers at age 14 years, 43% remained occasional smokers; 54% became regular smokers and 3% became non-smokers. All of those who were regular smokers at age 14 years remained as regular smokers at age 16 years.

3. Estimated Response Vectors. These show estimates of the probability that a subject who is in a given latent state will be classified in a given observed state. These probabilities thus provide measures of the reporting accuracy for different observed classifications. The estimates show clear evidence that reporting accuracy varied markedly conditional on the subject's latent status. Of those who were non-smokers, 96% were accurately classified by the report data and only 4% gave false positive reports. For those who were

regular smokers, reporting accuracy was also high with an estimated 95% of regular smokers being correctly identified by report data. However, the reporting of occasional smoking was highly imprecise and the estimates suggest that of those who were occasional smokers only 57% were accurately classified and 42% were false negatives who were incorrectly classified as non-smokers.

INSERT TABLE 3. HERE

Replication of Results Using Self Report Data

The preceding analysis has been based on smoking behaviors classified on the basis of combined parental and self report data. However, it has been suggested to us that model conclusions could depend on the methods by which smoking behaviors were assessed. To address this issue, the analysis was rerun using classifications of smoking behaviors based on self report only. Comparisons of the models for combined parental and self report data with the self report only data led to the following conclusions.

1. Both analyses led to the conclusion that the best fitting and most parsimonious model was one in which the latent transition process varied with age but response errors did not vary with age. The self report analysis yield a log likelihood chi square of 24.06 (df = 54; p>.99) for this model.

2. Both analyses yielded generally similar parameters and substantive interpretations. To test the similarity of parameters across models, tests of differences between the same estimate derived from different methods were conducted using pooled standard error estimates and “t” statistics. This

showed that in no cases were the parameters for the combined parent/self report data significantly different from the corresponding parameters based on self report alone. This result implies that the parameters of the model for self report data alone were within the limits of sampling error shown for the parameters in Table 3.

Collectively these results suggest that the analysis was robust to the choice of measurement method to the extent that the use of different measurement methods led to: a) the same model choice; b) similar model interpretation and c) parameters that were not significantly different across models.

Comparison of Observed and Latent Transition Processes

At this point it is useful to compare the estimated latent transition matrices in Table 3 with the corresponding observed transition matrices in Table 1. It is evident from this inspection that both sets of matrices have some features that are in common and some features that are different. The feature common to both sets of matrices is that both suggest that with the passage of time there is a clear tendency for the rates of transition to smoking behaviors to increase. However, the two sets of matrices differ in their representation of this process. The largest difference is that whilst the observed transition matrices suggest that occasional smokers often made transitions to being non-smokers, the latent Markov estimates suggest that this transition was relatively infrequent with only small numbers of occasional smokers returning to being non-smokers. The reasons for this very clear difference in observed and latent transition matrices is evident from the response vector given in Table 3. This shows that the reporting of smoking behavior by occasional smokers was highly imprecise with 42% of occasional smokers describing themselves as non-smokers. This tendency for occasional smokers to under report their smoking status led to a situation in which the observed data produced a serious over-estimation of the rates of transition from being an occasional smoker to a non-smoker. The model estimates suggest that most of the apparent return to non-smoking from being an occasional smoker was due to errors of measurement in the reporting of smoking behaviors rather than to genuine remission or cessation of smoking behaviors by occasional smokers.

Discussion

In this paper we have used methods of latent Markov analysis to describe the development of cigarette smoking behaviors from the age of ten to the age of 16 years**.** The aims of this analysis were to estimate the rates of transition between non-smoking, occasional smoking and regular smoking taking into account errors of measurement in the reporting of smoking behaviors. The results of this analysis have two sets of implications for the study of smoking behaviors. These implications relate to: a) the process by which young people make transitions to smoking behaviors; b) the extent to which observed report data were afflicted by errors of measurement and the consequences of these errors. These issues are discussed below.

The Process of Transition to Smoking Behaviors

The parameters of the latent Markov model suggest that when due allowance was made for errors of measurement in report data, the transition to smoking behavior was characterized by two features. First the development of smoking behaviors appears to be a progressive and largely one way process. In particular, the parameters of the latent transition matrix suggest that once young people had become occasional smokers, it was unlikely that they would return to being non-smokers and, similarly, once young people had become regular smokers it was unlikely that they would return to being occasional smokers or non-smokers. These results suggest a process in which the individual transitions through various stages of smoking were, to a substantial extent, irreversible so that once an individual had graduated to a given stage of smoking behavior it was unlikely that he/she would return to an earlier stage of this process.

A second feature of this process was that transitions to smoking behaviors showed a clear tendency to accelerate with age. For example, during the interval from 10 to 12 years less than 10% of those who were non-smokers at 10 became smokers at 12. Whereas during the interval from 14 to 16 years over a third of non-smokers at age 14 became smokers at age 16 years. There were similar trends to the transition from occasional smoking to regular smoking to accelerate with age.

These features of the development of smoking as being both progressive and accelerating with age have two implications for prevention programs. First, these results clearly suggest the importance of preventing young people from making transitions from being non-smokers to occasional smokers since the findings suggest that once young people have become occasional smokers it is unlikely that they will return to being non-smokers. Secondly, the results suggest the importance of targeting interventions at young people during the period from 14 to 16 years since this is the time at which there is an increasing rate of transition to smoking behaviors. These conclusions suggest the need for smoking prevention programs which begin relatively early in childhood and extend into adolescence with these programs being directed at different stages of the transition to smoking. The results suggest that, in early and middle childhood, the aims of effective programs should be to reduce the likelihood that young people will experiment with or occasionally use cigarettes but that the emphasis of these programs needs to change in adolescence to address the clear increase in rates of usage amongst those over 14 years.

These conclusions fit well with conclusions that we have drawn on the basis of a rather different analysis of these data. In particular in a previous study (Fergusson et al., 1994) we have found that two factors were major predictors of smoking by the age of 16 years: early onset of smoking behaviors and peer affiliations in adolescence. Collectively, these results suggest that optimal smoking prevention programs are likely to be those which are continued throughout childhood with the emphasis of these programs varying in an age appropriate way. During early and middle childhood the focus of such programs should be upon educational and other methods which deter children from smoking experimentation which appears to be the first step on the road to tobacco addiction. However, in adolescence such educational efforts should be supplemented by programs which address the effects of peer affiliations and peer pressure on rates of transition to occasional or regular smoking.

It should be noted that in this analysis we have examined the process of transition to regular weekly smoking. However, beyond this stage there is clearly a final stage in the development of cigarette smoking: that of becoming an addicted daily smoker. This transition has not been examined in this study, as the number of addicted daily smokers before the age of 16 years was too small for useful analysis. These results clearly suggest that the process of transition to regular addicted smoking is generally after the age of 16 in late adolescence and early adulthood. In future studies of this cohort we hope to study the transition to smoking behaviors up to the age of 18 years.

Implications for Errors of Measurement in Reports of Smoking Behavior

In addition to providing an account of the processes by which subjects made transitions from being non-smokers to smokers, the parameters of the latent Markov model can also be used to assess the accuracy of report data. In particular the estimates from the response vector show the estimated probabilities that a subject who was in a given latent state would report smoking behaviors (given the model assumptions). These estimates showed that reporting accuracy varied markedly with the subject's latent status. The estimates suggested that individuals who were non-smokers or regular smokers generally reported their smoking behavior with high accuracy with over 95% of non-smokers or regular smokers correctly reporting their smoking behaviors. However, the reporting accuracy of occasional smoking was poor and estimates suggested that in the region of 40% of occasional smokers made false negative responses which resulted in them being classified as non-smokers. There may be a number of reasons for the poor reporting accuracy of occasional smokers. However, an important reason may be that this group has an ambiguous status: its members are neither committed non-smokers nor committed smokers. Under such conditions of status ambiguity it may be that occasional smokers feel that they have considerable latitude to describe themselves as smokers or non-smokers depending on the way they feel at the time of questioning. This would result in those who are occasional smokers tending to produce false negative reports in which they were classified as non-smokers.

It has been pointed out to us that the latitude that occasional smokers have to under-report smoking may have been increased in the present study as a result of the fact that no exact timeframe was used to assess smoking behaviors. This could have given occasional smokers the latitude to implicitly truncate the timeframe over which they reported smoking and thus represent themselves as non-smokers. This possibility would clearly not have applied to non-smokers or regular smokers. These conclusions may suggest that one way of improving the reporting accuracy of occasional smoking may be to specify an exact timeframe for the reporting of smoking behaviors. Nonetheless, the very large reporting errors associated with occasional smoking status suggest that even if an exact timeframe were to be specified it is likely that reports of occasional smoking would be still subject to substantial response error.

These findings have two implications for the measurement of smoking behaviors during childhood. First, they suggest that attempts to improve measurement accuracy need to focus on methods which lead occasional smokers to report their behaviors more accurately. Here it is notable that it has often been claimed that the use of biochemical methods may be useful in improving reporting accuracy of smoking behaviors (Gillies et al., 1982; Luepker et al., 1981). However, these methods are likely to be relatively ineffective as a means of detecting occasional or experimental smokers suggesting that whilst biochemical methods may provide a useful means of validating reports of regular smoking these methods are unlikely to result in a substantial improvement in the accuracy of measurement of occasional smoking. Second, given that regular smoking appears to be reported with relatively high accuracy, the results here may suggest that the most robust way of measuring smoking behaviors in childhood may be to contrast regular smokers with other children rather than to compare smokers with non-smokers.

The findings of the analysis also made it possible to explore the implications of reporting errors for substantive conclusions about the process of transition to smoking behaviors. Comparisons of the observed data with the latent transition matrices revealed a number of similarities and differences between the processes of transition to smoking before and after correction for measurement errors. The largest difference was that whilst the observed data suggested that occasional smokers frequently made transitions back to being non-smokers, the latent model suggested the opposite conclusion: once children became occasional smokers return to non-smoking was unlikely. The reasons for these marked differences in the model estimates lie with the reporting accuracy of occasional smoking behaviors. In particular, since many occasional smokers gave false negative accounts of smoking behaviors these false negatives conveyed the misleading impression that occasional smokers were showing remission of behavior when, in fact, the latent estimates suggest that most of this apparent remission was due to measurement error rather than to genuine behavioral change.

Concluding Comment

Finally we wish to venture some brief comment on the realism of the model developed in this paper. The use of latent variable modelling methods to represent measurement errors in studies of substance use behaviors has been controversial with some authors claiming that such methods may resolve the problems of errors in report data (eg Newcomb & Bentler, 1988), whereas others have claimed that modelling methods may lack realism and may be misleading (Baumrind, 1983; Martin, 1982). We prefer to hold the middle ground in this debate and believe that whilst latent variable models may be subject to potential weakness and fallibility and should be treated with caution, these methods may also provide useful insights into both the extent and consequences of measurement errors in report data. In the present case the analysis suggests that the observed data were well described by a first order latent Markov model in which: a) rates of transition varied in an age related way; b) errors of measurement depended on the subject's latent status. In addition to fitting the data well this model also produced conclusions that accorded well with both theoretical and commonsense expectations about the development of smoking behaviors and sources of reporting errors. Specifically the model suggested the presence of a developmental process that was progressive and accelerated with time and that errors of measurement were most marked for those who were occasional smokers and had an ambiguous status. To the extent that the model fitted the data well and produced results which were in accord with theoretical and commonsense expectations, there is clearly some evidence for the realism of the model assumptions as at least a first approximation to an account of the way in which children make transitions to smoking behaviors during adolescence. At the same time, the fact that the proposed model fitted the data well and also produced results that were in accord with theoretical and commonsense expectations does not preclude the possibility that alternative models cannot be proposed to explain the data and its properties.

Author Notes

This research was funded by grants from the Health Research Council of New Zealand and the National Child Health Research Foundation.

References

Baumrind, D. (1983). Specious causal attributions in the social sciences: The reformulated stepping-stone theory of heroin as an exemplar. Journal of Personality and Social Psychology, 45, 1289-1298.

Berk, R.A. (1983). An introduction to sample selection bias in sociological data. American Sociological Review, 48, 386-398.

Chassin, L., & Presson, C.C. (1990). The natural history of cigarette smoking: predicting young-adult smoking outcomes from adolescent smoking patterns. Health Psychology, 9, 701-716.

Conrad, K.M., Flay, B.R., & Hill, D. (1992). Why children start smoking cigarettes: predictors on onset. British Journal of Addiction, 87, 1711-1724.

Escobedo, L.G., Marcus, S.E., Holtzman, D., & Giovino, G.A. (1993). Sports participation, age at smoking initiation, and the risk of smoking among US high school students. Journal of the American Medical Association, 269, 1391-1395.

Fergusson, D.M., & Horwood, L.J. (1989). A latent class model of smoking experimentation in children. Journal of Child Psychology and Psychiatry, 30, 761-773.

Fergusson, D.M., Horwood, L.J., Shannon, F.T., & Lawton, J.M. (1989). The Christchurch Child Development Study: A review of epidemiological findings. Paediatric and Perinatal Epidemiology, 3, 278-301.

Fergusson, D.M., Lynskey, M.T., & Horwood, L.J. (1994). The role of peer influences in the development of adolescent cigarette smoking. (Submitted for publication).

Gillies, P. (1985). Accuracy in the measurement of the prevalence of smoking in young people. Health Education Journal, 44, 36-38.

Gillies, P.A., Wilcox, B., Coates, C., Kristmundsdottir, F., & Reid, D.T. (1982). Use of objective measurement in the validation of self-reported smoking in children aged 10 and 11 years: saliva thiocyanate. Journal of Epidemiology and Community Health, 36, 205-208.

Goodman, L.A. (1974). The analysis of systems of qualitative variables when some of the variables are unobservable. Part I - A modified latent structure approach. American Journal of Sociology, 79, 1179-1259.

Holland, W.W., & Fitzsimmons, B. (1991). Smoking in children. Archives of Disease in Childhood, 66, 1269-1270.

Kandel, D.B. (1980). Drug and drinking behavior among youth. Annual Review of Sociology, 6, 235-285.

Krohn, M.D., Massey, J.L., Skinner, W.F., & Lauer, R.M. (1983). Social bonding theory and adolescent cigarette smoking: A longitudinal analysis. Journal of Health and Social Behavior, 24, 337-349.

Langeheine, R., & Van de Pol, F. (1990). A unifying framework for Markov modelling in discrete space and discrete time. Sociological Methods and Research, 18, 416-441.

Luepker, R.V., Pechaeck, T.F., Murray, D.M., Johnson, C.A., Hurd, P.D., & Jacobs, D.R. (1981). Saliva thiocyanate: a chemical indictor of cigarette smoking in adolescents. American Journal of Public Health, 71, 1320-1324.

McKennell, A.C. (1980). Bias in the reported incidence of smoking by children. International Journal of Epidemiology, 9, 167-177.

McNeil, A.D., Jarvis, M.T., West, R., Russell, M.A.H., & Bryant, A. (1987). Saliva cotinine as an indicator of cigarette smoking in adolescents. British Journal of Addiction, 82, 1355-1360.

Martin, J.A. (1982). Application of structural modelling with latent variables to adolescent drug use: A reply to Huba, Wingard and Bentler. Journal of Personality and Social Psychology, 43, 598-603.

Miller, S.K., & Slap, G. (1989). Adolescent smoking: A review of prevalence and prevention. Journal of Adolescent Health Care, 10, 129-135.

Moncher, M.S., Holden, G.W., & Schinke, S.P. (1991). Psychosocial correlates of adolescent substance use: A review of current etiological constructs. The International Journal of the Addictions, 26, 377-414.

Murray, D.M., & Perry, C.L. (1987). The measurement of substance use among adolescents: When is the "bogus pipeline" method needed? Addictive Behaviors, 12, 225-233.

Newcomb, M.D., & Bentler, P.M. (1988). Consequences of adolescent drug use: impact on the lives of young adults. Beverly Hills: Sage.

Stacy, A.W., Flay, B.R., Sussman, S., Brown, K.S., Santi, S., & Best, J.A. (1990). Validity of alternative self-report indices of smoking among adolescents. Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2, 442-446.

Taioli, E., & Wynder, E.L. (1991). Effect of age at which smoking begins on frequency of smoking in adulthood. New England Journal of Medicine, 325, 968-969.

Van de Pol, F., & de Leeuw, J. (1986). A latent Markov model to correct for measurement error. Sociological Methods and Research, 15, 118-141.

Van de Pol, F., & Langeheine, R. (1990). Mixed Markov latent class models. In C.C. Clogg (Ed.), Sociological methodology (pp. 213-247). Oxford: Blackwell.

Van de Pol, F., Langeheine, R., & de Jong, W. (1991). PANMARK user manual: panel analysis using Markov chains. Version 2.2. Voorburg: Netherlands Central Bureau of Statistics.

Wiggins, L.M. (1973). Panel analysis, latent probability models for attitude and behavior processes. Amsterdam: Elsevier.

Table 1: Observed probabilities of transition between smoking states (number of subjects making transitions are shown in parentheses)

	Non-Smoker		Occasional Smoker		Regular Smoker	
10 - 12 Years						
Non-Smoker	.896	(768)	.098	(84)	.006	(5)
Occasional Smoker	.563	(54)	.406	(39)	.031	(3)
Regular Smoker	.250	(1)	.500	(2)	.250	(1)
12 - 14 Years						
Non-Smoker	.825	(679)	.143	(118)	.032	(26)
Occasional Smoker	.472	(59)	.352	(44)	.176	(22)
Regular Smoker	.222	(2)	.111	(1)	.667	(6)
14 - 16 Years						
Non-Smoker	.666	(493)	.203	(150)	.131	(97)
Occasional Smoker	.282	(46)	.264	(43)	.454	(74)
Regular Smoker	.037	(2)	.056	(3)	.907	(49)

Table 2: Comparison of goodness of fit of alternative models

	Log Likelihood Ratio χ^2	d.f.	p
Model Goodness of Fit			
Model 1: (Constant transition matrix, constant response probabilities)	158.9	66	<.0001
Model 2: (Time dependent transition matrices, constant response probabilities)	31.9	54	>.99
Model 3: (Time dependent transition matrices, time dependent response probabilities)	26.3	48	>.99
Tests of Improvement in Fit			
Model 2 vs Model 1	127.0	12	<.0001
Model 3 vs Model 2	5.6	6	>.25

Table 3: Estimated parameters (standard errors) for Model 2

		Non-Smoker	Occasional Smoker	Regular Smoker
Initial (10 Year) Probabilities				
		.863 (.035)	.134 (.034)	.003[a] (.004)
Latent Transition Probabilities				
10 - 12 Years		Non-Smoker	Occasional Smoker	Regular Smoker
	Non-Smoker	.927 (.028)	.070 (.028)	.003[a] (.004)
	Occasional Smoker	.000[b] (-)	.977 (.040)	.023[a] (.040)
	Regular Smoker	.000[b] (-)	.000[b] (-)	1.000[b] (-)
12 - 14 Years		Non-Smoker	Occasional Smoker	Regular Smoker
	Non-Smoker	.861 (.031)	.124 (.032)	.015[a] (.010)
	Occasional Smoker	.067[a] (.093)	.744 (.092)	.189 (.047)
	Regular Smoker	.000[b] (-)	.000[b] (-)	1.000[b] (-)
14 - 16 Years		Non-Smoker	Occasional Smoker	Regular Smoker
	Non-Smoker	.615 (.044)	.291 (.049)	.094 (.025)
	Occasional Smoker	.035[a] (.085)	.425 (.118)	.539 (.067)
	Regular Smoker	.000[b] (-)	.000[b] (-)	1.000[b] (-)
Response Probabilities[c]			Observed Status	
		Non-Smoker	Occasional Smoker	Regular Smoker
	Non-Smoker	.964 (.011)	.035 (.011)	.001[a] (.001)
Latent Status	Occasional Smoker	.419 (.076)	.574 (.068)	.007 (.021)
	Regular Smoker	.000[b] (-)	.046[a] (.074)	.954 (.074)

[a] Parameter not significantly different from zero (p>.05)
[b] Parameter converged to zero or one during estimation - standard errors cannot be calculated[c] Probabilities of observed case status conditional on latent case status